Remarks

Applicant is amending the claims to include the words "speaker-dependent and text-dependent" when reciting grammars. Additionally, hyphens are being added to claims that already recited "speaker dependence" to be consistent with the hyphenation of "speaker-dependence" in the other claims. Additionally, certain antecedent basis issues have been addressed.

In the sixth and seventh paragraphs of the Official Action, Examiner rejected claims 1-3, 11-13 and 17-21, and also rejects claims 14-16 and 20-22 under 35 U.S.C. 102(e) as being anticipated by U.S. Patent 5,897,616 issued to Kanevsky ("Kanevsky"). Claims 1, 11 and 17 on which claims 2-3, 12-16 and 18-22 depend, are being amended merely to define the invention more specifically.

The claims recite the extraction and comparison of at least one speaker-dependent, text-dependent grammar. It is noted that grammars were defined as speaker-dependent and text-dependent in the specification, and argued as such in the response to the prior official action and so this amendment does not narrow the scope of the claims in any fashion. As noted previously, "A grammar describes how a particular phrase is spoken by a particular person".

Kanevsky does not use grammars. Kanevsky addresses recognizing (e.g. decoding) the text of the word spoken through automatic speech recognition (ASR) and then attempts to match the recognized text to a set of stored text. The system then performs a challenge/response test, whereby the user is requested to answer preferably several questions that only the user would know, and the user's response is recognized again using ASR and compared to the answers stored in the server. There is no mention of speaker dependence or text dependence for the ASR as claimed. The problem with using names is that ASR tends not to work well in the absence of known pronunciation patterns. Because names need not follow a universal set of pronunciation rules, ASR is not well suited for them, even though Kanevsky uses them. Kanevsky attempts to get around this problem by asking several challenge/response questions, which can improve accuracy of the recognition but would annoy the user.

It is noted that Kanevsky also addresses the use of acoustic models of the speaker to further verify identification of the user, but those acoustic models are described throughout Kanevsky as text-independent. Acoustic models describe information about the user's voice, but not how the user speaks a particular phrase,

which is what a grammar does. Thus, the acoustic models are not the same as grammars. The fact that Kanevsky notes that these are text independent (Col. 5, line 36; Col. 7, line 3, Col. 7, line 64) means that Kanevsky teaches precisely away from the claimed invention of text-dependence.

Thus, claims 1-3, and 11-22 are patentably distinguishable over Kanevsky.

In paragraphs 8 and 9 of the Official Action, Examiner rejects claims 4-10 under 35 U.S.C. 103(a) as being unpatentable over Kanevsky. Claim 1, on which claims 4-10 depend, is being amended merely to define the invention more specifically.

Examiner relies on the 102 rejection above for this rejection, and as described above, claim 1, from which claims 4-10 ultimately depend, is patentably distinguishable over Kanevsky. In addition, Examiner's reasoning for making the modification, less memory required, could not be located by Applicants anywhere in Kanevsky, and the source of examiner's reasoning, that for a small group of users, the recognizer would need a small memory, isn't at all clear. Kanevsky is geared towards a large group of users (Col. 2, line 15). The purpose behind

Kanevsky is to limit the number of users among which the identity is to be discerned, which is a problem with a large group of users, not a small group. Furthermore, it isn't clear how the addition of the second recognizer would cause the system to require less memory or other storage, when the second recognizer would have to operate in memory, increasing the memory requirements, not decreasing them as claimed. Thus, claims 4-10 are also patentably distinguishable over Kanevsky.

Version Showing Changes

1. (amended) A system for identifying a selected user from a first plurality of users, the system comprising:

a first grammar extractor having a first input operatively coupled to receive an identifier of one of the plurality of users, and a second input operatively coupled to receive a first utterance from the one of the first plurality of users uttered during a first session, the first grammar extractor for extracting a first speaker-dependent, text-dependent grammar from the first utterance received at the first grammar extractor first input and for providing at an output the first speaker-dependent, text-dependent grammar and the corresponding identifier received at the first input;

a grammar storage having an input/output coupled to the first grammar extractor output, for receiving the first speaker-dependent, text-dependent grammar and identifier for each of the plurality of users and storing the first speaker-dependent, text-dependent grammar responsive to the identifier, and for providing at the input/output one of said first speaker-dependent, text-dependent grammars

corresponding to an identifier responsive to receipt of said identifier at the grammar storage input/output;

a second grammar extractor having an input operatively coupled to receive a second utterance from the selected user uttered during a second session different from the first session, the second grammar extractor for extracting and providing at an output a second speaker-dependent, text-dependent grammar responsive to the second utterance received at the second grammar extractor input; and

a first recognizer having a first input coupled to the grammar storage input/output, and a second input coupled to the second grammar extractor output, the first recognizer for identifying a match between a set of a plurality of the first speaker-dependent, text-dependent grammars stored in the grammar storage and the second speaker-dependent, text-dependent grammar received at the second first recognizer input, and for providing at an output coupled to an apparatus output, the identifier of the user corresponding to a one of the first speaker-dependent, text-dependent grammar in the grammar storage most closely matching the second speaker-dependent, text-dependent grammar received at the first recognizer second input.

2. The system of claim 1, wherein the first utterance comprises a password of the one of the plurality of users, and the second utterance comprises a password of the user.

3. The system of claim 1, wherein the first grammar extractor is the second grammar extractor.

4. The system of claim 1, additionally comprising:

A second recognizer having an input operatively coupled to receive a third utterance uttered during the second session, the second recognizer for, responsive to the third utterance, identifying a second plurality of users from the first plurality of users as belonging in the set of users.

5. The system of claim 4, wherein the third utterance comprises a name of the user.

6. The system of claim 5, wherein the second recognizer identifies the second plurality of users responsive to the third utterance by recognizing the third utterance and comparing the recognized third utterance with a list of user identifiers of the first plurality of users.

7. The system of claim 6, additionally comprising:

a first voiceprint extractor having a first input operatively coupled to receive an identifier of one of the

first plurality of users and a second input operatively coupled to receive a fourth utterance from the one of the first plurality of users uttered during the first session, the first voiceprint extractor for creating a voiceprint responsive to the fourth utterance and for providing the voiceprint and the identifier of the user at an output;

a voiceprint storage having an input/output coupled to the first voiceprint extractor output, the voiceprint storage for storing, for each of the first plurality of users, the voiceprint received at the voiceprint storage input/output associated with the identifier of the user, and for providing a first voiceprint at the input/output responsive to a request for the voiceprint comprising the identifier of the user corresponding to the voiceprint received at the voiceprint storage input/output;

a second voiceprint extractor having an input coupled to receive a fifth utterance uttered by the selected user during the second session, the second voiceprint extractor for extracting and providing at an output a second voiceprint responsive to the fifth utterance; and

a verifier having a input/output coupled to the voiceprint storage input/output, a first input coupled to the second voiceprint extractor output for receiving the

second voiceprint, and a second input coupled to the first recognizer output, and an output coupled to the apparatus output, the verifier for providing at the input/output an identifier corresponding to the identifier received at the second input and receiving at the input/output one of the first voiceprints, said first voiceprint corresponding to the identifier provided at the input/output, the verifier additionally for comparing the first voiceprint received at the verifier input/output with the voiceprint received at the first verifier input and for signaling at an output coupled to the apparatus output responsive to said first voiceprint and said second voiceprint matching within an acceptable tolerance level.

8. The system of claim 7 wherein:

each fourth utterance comprises a sixth utterance uttered by one of the plurality of users during the first session and the first utterance; and

the fifth utterance comprises the second utterance and the third utterance.

9. The system of claim 8, wherein the sixth utterance comprises a name of one of the first plurality of users.

10. The system of claim 7 wherein the second voiceprint extractor is the first voiceprint extractor.

11. (amended) A method of identifying a caller as a user of a computer system, the method comprising:

receiving a first utterance;

extracting a speaker-dependent, text-dependent grammar from the first utterance;

comparing the speaker-dependent, text-dependent grammar extracted with a set of speaker-dependent, text-dependent grammars, each speaker-dependent, text-dependent grammar in the set of speaker-dependent, text-dependent grammars corresponding to a user;

responsive to the comparing the speaker-dependent, text-dependent grammar step, identifying a set of at least one user, the number of users in set of at least one user smaller than the number of users corresponding to the speaker-dependent, text-dependent grammars in the set of speaker-dependent, text-dependent grammars; and

extracting a voiceprint from the first utterance;

comparing the voiceprint extracted with a voiceprint for each user in the set of at least one user; and

identifying the user responsive to the comparing the voiceprint step.

12. The method of claim 11, wherein the number of users in the set of at least one user is one.

13. (amended) The method of claim 11, additionally comprising:

receiving a second utterance from the caller;

recognizing the second utterance; and

identifying the set of speaker-dependent, text-dependent grammars responsive to the recognizing the second utterance step.

14. The method of claim 13, wherein the recognizing step comprises speaker independent voice recognition of the second utterance.

15. (amended) The method of claim 13, wherein the recognizing step comprises [speaker dependent] speaker-dependent voice recognition of the second utterance.

16. The method of claim 13, wherein the extracting the voiceprint step comprises extracting the voiceprint from the first utterance and the second utterance.

17. (amended) A computer program product comprising a computer useable medium having computer readable program code embodied therein for identifying a caller as a user of a computer system, the computer program product comprising:

computer readable program code devices configured to cause a computer to receive a first utterance;

computer readable program code devices configured to cause a computer to extract a speaker-dependent, text-dependent grammar from the first utterance;

computer readable program code devices configured to cause a computer to compare the speaker-dependent, text-dependent grammar extracted with a set of speaker-dependent, text-dependent grammars, each speaker-dependent, text-dependent grammar in the set of speaker-dependent, text-dependent grammars corresponding to a user;

computer readable program code devices configured to cause a computer to, responsive to the computer readable program code devices configured to cause a computer to compare the speaker-dependent, text-dependent grammar, identify a set of at least one user, the number of users in set of at least one user smaller than the number of users corresponding to the speaker-dependent, text-dependent grammars in the set of speaker-dependent, text-dependent grammars;

computer readable program code devices configured to cause a computer to extract a voiceprint from the first utterance;

computer readable program code devices configured to cause a computer to compare the voiceprint extracted with a voiceprint for each user in the set of at least one user; and

computer readable program code devices configured to cause a computer to identify the user responsive to the comparing the voiceprint step.

18. The computer program product of claim 17, wherein the number of users in the set of at least one user is one.

19. (amended) The computer program product of claim 17, additionally comprising:

computer readable program code devices configured to cause a computer to receive a second utterance from the caller;

computer readable program code devices configured to cause a computer to recognize the second utterance; and

computer readable program code devices configured to cause a computer to identify the set of speaker-dependent, text-dependent grammars responsive to the recognizing the second utterance step.

20. The computer program product of claim 19, wherein the computer readable program code devices configured to

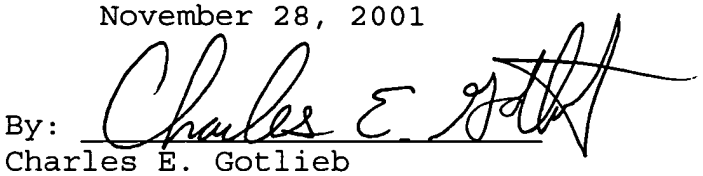
cause a computer to recognize comprise computer readable program code devices configured to cause a computer to perform speaker independent voice recognition of the second utterance.

21. (amended) The computer program product of claim 19, wherein the computer readable program code devices configured to cause a computer to recognize comprise computer readable program code devices configured to cause a computer to perform [speaker dependent] speaker-dependent voice recognition of the second utterance.

22. The computer program product of claim 19, wherein the computer readable program code devices configured to cause a computer to extract the voiceprint comprise computer readable program code devices configured to cause a computer to extract the voiceprint from the first utterance and the second utterance.

Thus, claims 1-22 are patentably distinguishable over the cited references. Favorable action is solicited.

Respectfully submitted,
November 28, 2001

By: 

Charles E. Gotlieb
Registration No. 38,164
Innovation Partners
540 University Ave., Suite 300
Palo Alto, CA 94301
(650) 328-0100